Lingdex.org: Leveraging LLMs to Structure and Explore Linguistic Olympiad Puzzles for Learning and Teaching Linguistics

Jonathan Sakunkoo
University of Oxford
jonathan@sakunkoo.com

Annabella Sakunkoo
Stanford University OHS
apianist@ohs.stanford.edu

Abstract

Linguistics Olympiad puzzles provide a valuable but underutilized resource for teaching linguistic reasoning, typology, and cross-cultural understanding. Many of these problems feature endangered and low-resource languages and thus offer a rare opportunity to integrate linguistic diversity into education at a time when over 40% of the world's languages face extinction. This paper presents Lingdex, a novel web-based platform that leverages large language models (LLMs) to classify, organize, and enliven linguistics Olympiad problems across various linguistic categories such as syntax, morphology, semantics, phonology, and language families. By applying NLP techniques to the multilingual and multicultural corpora of linguistic puzzles drawn from international and national Olympiads, Lingdex supports language and linguistics education, problem-based learning, and curriculum development. The visual, interactive platform also includes problems based on endangered and rare languages to raise awareness of and interest in linguistic diversity. We present results from a user study that shows increased learner interest and appreciation for global linguistic richness.

1 Introduction

How do you spell and pronounce your name in Māori? To what English word do you think the following word in Māori correspond: *pirinihehe*? Have you heard of Mongo, a Bantu language?

Linguistics Olympiad puzzles challenge students to analyze language data and discover patterns from unfamiliar languages. The puzzles can teach linguistic reasoning, typology, problem-solving, and cross-linguistic cultural awareness. These puzzles often feature endangered or low-resourced languages and offer a unique opportunity to promote interest in linguistics and global linguistic diversity.

Although there are over 7,000 living languages in the world, over 40% are considered endangered. As languages die at a rate of one every two weeks, it is estimated that over half of the world's languages could disappear before the end of the century. Many languages lack formal documentation and face extinction within a generation without active efforts to preserve or revitalize them. Studying diverse languages is not only academically important, but also essential to preserve cultural knowledge and human cognitive diversity. Despite this urgency, linguistic and language education in classrooms typically centers on widely-spoken languages and mainstream linguistic phenomena.

Lingdex addresses this critical gap by providing structured, comprehensive, and interactive access to puzzles based on diverse, often endangered languages from linguistics Olympiad puzzles from the International Linguistics Olympiad and from many world regions in America, Asia, Europe, and Australia. By using LLMs to classify, organize, and enliven this fragmented and niche but highly educational content, the NLP-powered Lingdex transforms challenging and underutilized resources into a rich pedagogical tool.

This paper introduces Lingdex, outlines its LLM-based application, and demonstrates its potential educational impact. We show that the NLP-powered educational tool that integrates, structures, and visualizes puzzles from diverse languages can spark curiosity, strengthen analytical skills, and promote awareness of language endangerment.

2 Background

An increasing number of organizations are leveraging NLP in educational applications for teaching and assessment across domains such as language learning, mathematics, science, and programming (Sakunkoo and Sakunkoo, 2025a). For example, Rozovskaya (2024) emphasized multilingual low-resource NLP for language learning, and Siyan et al. (2024) has built an English-teaching chatbot with

Proceedings of Recent Advances in Natural Language Processing,pages 1053–1057
Varna, Sep 8–10, 2025
https://doi.org/10.26615/978-954-452-098-4-121

empathetic feedback. NLP has also been used in medical science education and writing evaluation (Klebanov and Madnani, 2020; Yaneva et al., 2024). However, its use in promoting linguistic diversity in education remains limited.

Linguistics Olympiads uniquely feature diverse, rare, and endangered languages. Prior work has

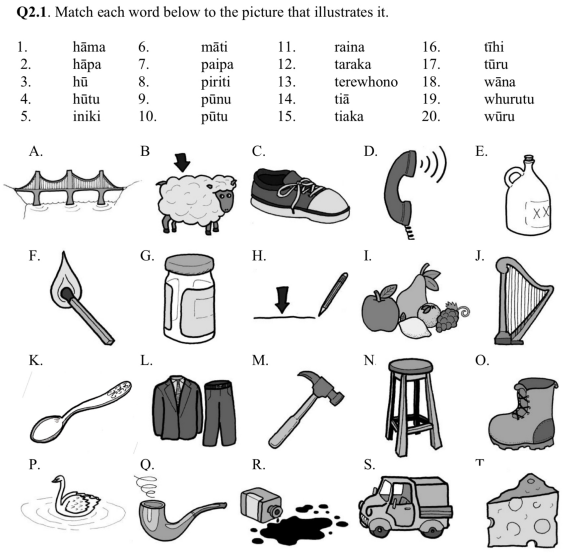

Figure 1: A sample Māori linguistic puzzle from UKLO

mostly focused on using NLP techniques to solve linguistic olympiad-style puzzles. Recently, Baddepudi et al. (2024) have attempted to apply deep-learning approaches to solving linguistic puzzles from small data. Other prior work also focused on enhancing the reasoning abilities of LLMs in solving complex linguistic puzzle tasks (Lin et al., 2023; Bean et al., 2024; Zhu et al., 2025).

Linguistic diversity is essential for preserving the full range of human knowledge, cultural expression, and cognitive perspectives in the world's languages and "understanding why and how languages differ tells us about the range of what is human" (Jurafsky, 2019). Rather than focusing on advancing LLMs' abilities to solve linguistics Olympiad puzzles, our work offers a novel NLP-powered tool to promote students' curiosity, interest, knowledge, and appreciation for linguistic analysis and diversity as well as understudied linguistic phenomena (Sakunkoo and Sakunkoo, 2025b). Lingdex uniquely applies LLMs to organize, classify, and enliven linguistic content rooted in global linguistic diversity, thus turning niche Olympiad problems into accessible, visual, and intelligent resources for students and teachers and supporting endangered language awareness.

3 Data and Methods

3.1 Data

Lingdex's dataset comes from the International Linguistics Olympiad (IOL) and National linguistics Olympiads, specifically NACLO (North America), UKLO (United Kingdom), OzCLO (Australia), APLO (Asia Pacific), and HKLO (Hong Kong). The dataset covers 100+ languages including endangered languages like Warlpiri, Paunaka, Dâw, and Tariana. It also includes isolates and lesser-known families such as Basque, which is believed to be the only remaining spoken descendant of the languages that were spoken before Proto-Indo-European speakers migrated into Europe, and Totonacan, a language isolate within Mesoamerica. Each puzzle contains data examples, analytic tasks, and cultural or linguistic context. They are valuable not only as linguistic challenges but also as mini-portraits of linguistic worlds.

3.2 Methods

Lingdex processes and classifies linguistics puzzles through a multi-step process that combines rule-based extraction and LLM-based classification.

Problem Extraction and Parsing We begin by collecting problems in several formats, including HTML, PDF, and scanned images, from national and international contests. Each contest has its own formatting conventions, so we created a custom adapter for each source. These adapters segment each year into rounds, and then into individual problems. For every problem, two main text components are extracted using regex: the problem statement and the solution explanation. Metadata fields such as problem title, author, year, round, and problem number are also parsed.

Classification with LLMs We compare and use GPT-4o-mini and LLama3.2 to classify each problem along these dimensions: language (e.g. Māori, Basque), language family (e.g. Austronesian, Bantu), linguistic topic (e.g. phonology, syntax, semantics), question format (e.g. translation, matching, reconstruction), theme (e.g. number systems, kinship terms, animal names), and location (approximate geographic location of speakers).

The LLMs are prompted with a glossary of definitions sourced from the UKLO website to standardize responses. Prompts are few-shot and constrained to return exactly four lines per problem,

each corresponding to one classification field. Results are saved in TSV format and partially manually verified and corrected in a spreadsheet by knowledgeable humans. As LLM results are unpredictable and typically are not exactly of the desired format, some further postprocessing is required before the responses can be extracted and fed into the search engine. Using a rule-based approach, we attempt to automatically convert improper responses into valid responses, but manual tagging is required in the event that it fails to do so.

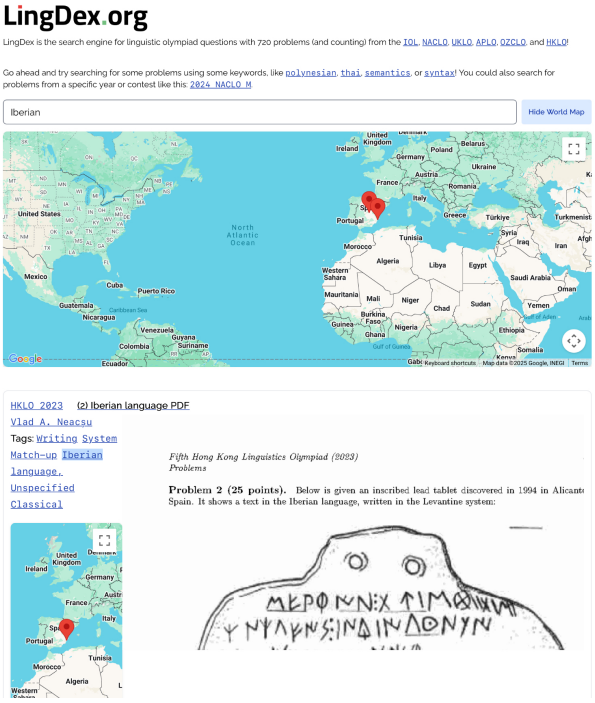

Figure 2: Lingdex transforms fragmented linguistic puzzles from around the world into NLP-powered system with interactive maps and hints for learning linguistics.

Visual Geographic Location Markers Another feature of Lingdex is the visual geographic location marker for each problem and its corresponding language: most of the problems are displayed along with an interactive world map with the featured language's approximate geographic location marked on it. This data is obtained by prompting the chosen LLM for the latitude and longitude in addition to the problem's classification. The LLM results tend to include extraneous and unwanted content, and regex text parsing was applied to clean the data (by filtering out non-numeric and non-punctuation characters). In the cases where the location could not be extracted, the map is not displayed. Additionally, Lingdex features a large world map that displays either the locations of all of the problems in Lingdex or, if there is an active search query, all the problems that fall under that query's category (e.g. all languages classified as "syntax").

Indexing and Search Engine All processed problems and metadata are indexed in Meilisearch to enable fast, faceted search. Users can filter problems by language, topic, country, year, or theme for custom lesson planning or linguistic exploration.

4 System Description

Following Nielsen's heuristics for UI design (Nielsen, 1994) and design principles that users prefer simplicity and familiarity, Lingdex is a user-friendly web tool offering searchable access by linguistic topic, language, language family, competition, and other keywords. It allows educators and students to search for and group puzzles into thematic lessons such as "Māori", "Austronesian", "number", "syntax", or "phonology". For usability,

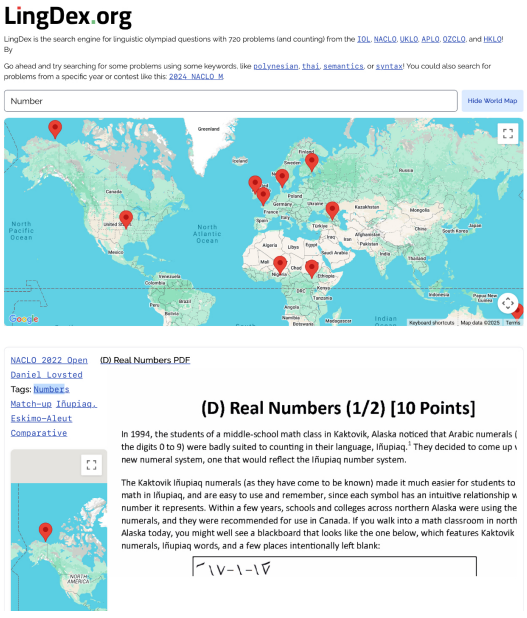

Figure 3: Sample Results Page after a search for "Number". It presents number problems in various languages such as Iñupiaq, with a map showing the location.

the system also generates thumbnail previews of both problems and solutions. These visual snippets allow users to quickly skim search results and select puzzles of interest. Lingdex also has an interactive world map that shows geographic locations of languages. It helps users explore puzzles by region and connect linguistic patterns to cultural and geographic contexts, while making learning more visual. Lingdex also integrates LLM capabilities to provide hints, which support user learning

by guiding problem-solving while encouraging independent reasoning and discovery. These hints function as information scent to help users stay engaged (Card et al., 2001; Sakunkoo, 2009).

5 Comparative Analysis of LLMs

This section critically analyzes strengths and limitations of two LLM models when applied to the domain of linguistics Olympiad problems, which is a unique and underexplored domain.

We evaluated GPT4o-mini and LLama3.2 on their ability to classify problems using UKLO's tagging as UKLO not only provides a glossary for relevant linguistic terms, but it is also the only competition to tag its problems by keywords and categories such as languages and subfields of linguistics such as semantics and phonology. Both models demonstrated high performance in tasks involving language and language family classification: GPT4o-mini achieved a high accuracy rate of 98% while Llama3.2 had 95% accuracy. However, when classifying problems by linguistic subfields such as phonology, morphology, syntax, and semantics, we observed notable divergence. GPT-4o-mini reached 72% accuracy, while LLama3.2 showed substantially lower performance at 36%. Importantly, both models struggled most with semantic problem classification, while they achieved high classification accuracy for number systems and writing systems. This suggests challenges in modeling deeper levels of abstract linguistic reasoning. By grounding our analysis in real problem classification tasks, we offer a foundation for future research into linguistically aware LLMs, especially in low-resource and educational settings.

6 User Study

We design a pilot study to assess whether Lingdex increases interest in linguistics and linguistic diversity, perceived competence, and satisfaction with its usefulness compared to being left to search for linguistics Olympiad problems across various sources independently. We hypothesize that Lingdex results in greater interest in linguistic diversity, perceived linguistic abilities, and satisfaction.

Twelve American users were recruited to use and evaluate Lingdex. Participants represented a mix of secondary and college students with varying levels of prior exposure to linguistics, ranging from complete beginners to those with experience in linguistic Olympiad-style problems. Participants engaged with Lingdex by searching for problems categorized by languages, language families, and linguistic features or clicking on the world map, each participant solving at least four problems. Before and after the study, students completed surveys measuring their interest in linguistics and linguistic diversity, perceived linguistic competence, and satisfaction with the usefulness of the application.

7 Findings and Conclusion

Lingdex users reported significantly higher engagement and learning satisfaction, compared to when they were instructed to train for linguistics Olympiad by finding and solving linguistics problems on their own through web searches and national linguistics resources. All participants reported increased interest in rare and endangered languages, greater awareness of linguistic diversity, and stronger motivation to learn more. Ten students described that they were excited by fun and interesting languages they had never known about. Eighty three percent said it helped them better understand linguistic concepts and typology and learn more efficiently and effectively. Every participant gave a high rating for the usefulness of visual maps, with eight users giving the highest rating. Ten participants reported feeling more competent and confident in solving linguistics puzzles.

Our results show that Lingdex makes linguistics Olympiad problems more accessible, engaging, and educational, especially those featuring endangered and rare languages. Powered by LLMs to organize and index problems across pedagogically meaningful categories, it supports educators, excites learners, and increases understanding, interest, and awareness of languages. As global language loss accelerates, innovative tools like Lingdex can enhance knowledge and promote appreciation for linguistic diversity. Future work includes adding multimodal resources such as sounds and videos, expanding the corpus, offering adaptive guidance, and localizing for diverse, multicultural learners.

Acknowledgments

Special thanks to Todd Krause, Chris Donlay, Patty Sakunkoo, and Jon Rawski for valuable teaching and guidance. We also thank USA International Linguistics Olympiad (IOL) team members and training hosts, NACLO organizers (especially Lori Levin), and IOL organizers and participants, whose inspiration and insights enriched this work.

References

Anavi Baddepudi, Emma Wang, and Ishan Khare. 2024. Minimal clues for maximal understanding: Solving linguistic puzzles with rnns, transformers, and llms. https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1244/final-projects/AnaviBaddepudiEmmaWangIshanKhare.pdf. Stanford CS224N Final Project.

Andrew Bean, Simi Hellsten, Harry Mayne, Jabez Magomere, Ethan A. Chi, Ryan Chi, Scott A. Hale, and Hannah Rose Kirk. 2024. Lingoly: a benchmark of olympiad-level linguistic reasoning puzzles in low-resource and extinct languages. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA. Curran Associates Inc.

Stuart K. Card, Peter Pirolli, Mija Van Der Wege, Julie B. Morrison, Robert W. Reeder, Pamela K. Schraedley, and Jenea Boshart. 2001. Information scent as a driver of web behavior graphs: results of a protocol analysis method for web usability. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, page 498–505, New York, NY, USA. Association for Computing Machinery.

Dan Jurafsky. 2019. The power of language: How words shape people, culture. https://news.stanford.edu/stories/2019/08/the-power-of-language-how-words-shape-people-culture. Interview featured in Stanford News.

Beata Beigman Klebanov and Nitin Madnani. 2020. Automated evaluation of writing—50 years and counting. *Computers and Composition*, 55:102543.

Zheng-Lin Lin, Chiao-Han Yen, Jia-Cheng Xu, Deborah Watty, and Shu-Kai Hsieh. 2023. Solving linguistic olympiad problems with tree-of-thought prompting. In *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, pages 262–269, Taipei City, Taiwan. The Association for Computational Linguistics and Chinese Language Processing (ACLCLP).

Jakob Nielsen. 1994. 10 usability heuristics for user interface design. https://www.nngroup.com/articles/ten-usability-heuristics/.

Alla Rozovskaya. 2024. Proceedings of the 19th workshop on innovative use of nlp for building educational applications. https://aclanthology.org/2024.bea-1.0.pdf. Keynote talk.

Annabella Sakunkoo and Jonathan Sakunkoo. 2025a. Name of thrones: How do LLMs rank student names in status hierarchies based on race and gender? In *Proceedings of the 20th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2025)*, pages 697–707, Vienna, Austria. Association for Computational Linguistics.

Jonathan Sakunkoo and Annabella Sakunkoo. 2025b. Lost and found: Computational quality assurance of crowdsourced knowledge on morphological defectivity in Wiktionary. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 998–1003, Vienna, Austria. Association for Computational Linguistics.

Patty Sakunkoo. 2009. Models of performance and behavior. Stanford HCI Group CS376 course presentation: https://hci.stanford.edu/courses/cs376/2009/lectures/2009-05-21-models/CS376_discussion_20090521-models.ppt.

Li Siyan, Teresa Shao, Julia Hirschberg, and Zhou Yu. 2024. Using adaptive empathetic responses for teaching English. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 34–53, Mexico City, Mexico. Association for Computational Linguistics.

Victoria Yaneva, King Yiu Suen, Le An Ha, Janet Mee, Milton Quranda, and Polina Harik. 2024. Automated scoring of clinical patient notes: Findings from the Kaggle competition and their translation into practice. In *Proceedings of the 19th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2024)*, pages 87–98, Mexico City, Mexico. Association for Computational Linguistics.

Hongpu Zhu, Yuqi Liang, Wenjing Xu, and Hongzhi Xu. 2025. Evaluating large language models for in-context learning of linguistic patterns in unseen low resource languages. In *Proceedings of the First Workshop on Language Models for Low-Resource Languages*, pages 414–426, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.